

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Merck & Co., Inc.
- (ii) TITLE OF INVENTION: CLONING AND IDENTIFICATION  
OF THE MOTILIN RECEPTOR
- (iii) NUMBER OF SEQUENCES: 15
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Merck & Co., Inc.
  - (B) STREET: P.O. Box 2000, 126 E. Lincoln Ave.
  - (C) CITY: Rahway
  - (D) STATE: NJ
  - (E) COUNTRY: USA
  - (F) ZIP: 07065-0900
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: Windows
  - (D) SOFTWARE: FastSEQ for Windows Version 2.0b
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 60/089,098
  - (B) FILING DATE: 12-JUN-1998
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Giesser, Joanne M
  - (B) REGISTRATION NUMBER: 32,838
  - (C) REFERENCE/DOCKET NUMBER: 20251 PCT
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 732-594-3046
  - (B) TELEFAX: 732-594-4720
  - (C) TELEX:

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3066 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTGAAATTAT	CTGGTCACTG	CCGGGCGCGG	TGGCTCACGC	CTGTAATCCC	AGCACTTTGG	60
GAGGTCGAGG	CGGGTGGACC	ACCTGGGGTC	AGGAGTTCGA	GACCAGGCTG	GCCAACATGG	120
CGAAACCCTG	ACTACACAAA	AAACACAAA	TTAGCCGGG	GCTTGGGCGC	TCCTGTGCTC	180
CCAGCTACTC	AGGAGGCTGA	GGTGGGAGGA	CTGCTTGAGC	CTGGGAGGTC	GAGGCTGCAG	240
TGAGCTGTGA	TCGCGCCACT	TAAACTCCAG	CCTGGACGAC	AGTGAGACCC	TGTCTCAAGA	300
AGAAAAAAG	AAAGAAAGAA	AGAAAAAAG	AAAAAAAAGA	AATTATTTGG	TCAATTATAT	360
GGTCAGCTCC	CTCCACCACT	CGCGAATTTA	CAGAAGAGGA	GAAGTGGGCT	GGGCGAGACC	420
AGGACTAGCC	CAAGATTACA	CAAGTTACTC	GGTTGTAGAG	CCAGGATTAG	ACAGGAGAGG	480
CTCTAGATTC	TGGTCTAGAC	TCCCCTCCTA	TTATTTAGCA	TTATGGCTTC	CTGAGGATTA	540
CCATGAGCCC	TCCTCCACCG	TCAAGCGGCA	GCTACCAGCC	ACCAGACCAG	ATCCCCTCGA	600
AGGTGCCCCG	AGTACACAGC	TGACAAAAGC	GCCCGTACAG	TGCTCAGTCC	TGTAACCAAA	660
GCTGTCTAGG	GTGCAGACAT	CGCTCACCGG	ACCGGGTAGG	GCTCGTGCGC	TAAGGGCGCC	720
GGGTATTCCA	GTTAGTGGAG	AGGGAAGCGC	CCTGGAAC TG	CATGGGCCCC	GGAAGGGCGC	780
CGGGAGCGGA	GCATGGCCCG	GCCGGGGCGG	GCCGCGGCCG	TGGGCGGAGA	CTGCGCGCAG	840
CTAGCTCGGG	AGCGCCTCGG	AGCCCACCCC	GCAGAGCCGC	TTCTCGCGCC	CCGCAGCGCA	900
GCGCAGCGCT	CCGCCGTCTG	ACCTGCCCGC	CCCGCAGCGT	GCGGGCTGGG	AAAGGAGGCG	960
CTCACCAGGA	GGGACCACGC	GCCAGGCTCC	CAGCCCGACC	CGGGACGCGG	CGGCCGCGCG	1020
GAGCACCCAT	GGGCAGCCCC	TGGAACGGCA	GCGACGGCCC	CGAGGGGGCG	CGGGAGCCGC	1080
CGTGGCCCGC	GCTGCCGCCT	TGCGACGAGC	GCCGCTGCTC	GCCCTTTCCC	CTGGGGGCGC	1140
TGGTGCCGGT	GACCGCTGTG	TGCCCTGTGC	TGTTTCGTCG	CGGGGTGAGC	GGCAACGTGG	1200
TGACCGTGAT	GCTGATCGGG	CGCTACCGGG	ACATGCGGAC	CACCACCAAC	TTGTACTCTG	1260
GCAGCATGGC	CGTGTCGGAC	CTACTCATCC	TGCTCGGGCT	GCCGTTCGAC	CTGTACCGCC	1320
TCTGGCGCTC	GCGGCCCTGG	GTGTTGCGGC	CGCTGCTCTG	CCGCCTGTCC	CTCTACGTGG	1380
GCGAGGGCTG	CACCTACGCC	ACGCTGCTGC	ACATGACCGC	GCTCAGCGTC	GAGCGCTACC	1440
TGGCCATCTG	CCGCCCGCTC	CGCGCCCGCG	TCTTGGTCAC	CCGGCGCCGC	GTCCGCGCGC	1500
TCATCTGGTG	GCTCTGGGCC	GTGGCGCTGC	TCTCTGCCCG	TCCCTTCTTG	TTCTTGGTGG	1560
GCGTCGAGCA	GGACCCCGGC	ATCTCCGTAG	TCCCGGGCCT	CAATGGCACC	GCGCGGATCG	1620
CCTCCTCGCC	TCTCGCCTCG	TGCGCCGCTC	TCTGGCTCTC	GCGGGCGCCA	CCGCCGTCCC	1680
CGCCGTCGGG	GCCCCGAGACC	GCGGAGGCCG	CGGCGCTGTT	CAGCCGCGAA	TGCCGGCCGA	1740
GCCCCGCGCA	GCTGGGCGCG	CTGCGTGTC	TGCTGTGGGT	CACCACCGCC	TACTTCTTCC	1800
TGCCCTTTCT	GTGCCTCAGC	ATCCTCTACG	GGCTCATCGG	GCGGGAGCTG	TGGAGCAGCC	1860
GGCGGCCGCT	GCGAGGCCCG	GCCGCCTCGG	GGCGGGAGAG	AGGCCACCGG	CAGACCGTCC	1920
GCGTCTGCG	TAAGTGGAGC	CGCCGTGGTT	CCAAAGACGC	CTGCCTGCAG	TCCGCCCCGC	1980
CGGGGACCGC	GCAAACGCTG	GGTCCCCCTC	CCCTGCTCGC	CCAGCTCTGG	GCGCCGCTTC	2040
CAGCTCCCTC	CTATTTTCGAT	TCCAGCCTCC	ACCCGCCGGT	ACTTCCCATC	CCCCGAGAAA	2100
ACCATGTCCT	GTCCCCCAGG	AGCTCTGGGG	GACCCACGGG	CGCTTTGAGG	GTGGGATCCC	2160
CGGATCCGAT	TCAGTAACCA	GCAGTGCTTT	TCCAGAGCCT	CTGAGACCAG	AAAGGAGAGT	2220
TGGTAATTCT	TAATCCAACC	ACCTGTTAGA	TGCCACAAAT	GAGGAGTCCT	CACAGTGCTC	2280
TTGAGAAGAC	GAGGGAGATT	TCATTAAGCT	AAAAATTTTT	ATTTAATGTT	AAGTGATGCT	2340
GAAGGCTAAA	GTAAACCTTG	CTCGTATCAA	AAAGTAAAGA	TTGTGCAGAC	CTGTTGTAGA	2400
ATTCTTTTCA	ACAGAGAACA	GAAAACCTGT	CTCCGAAGTG	GGTTTGTGGA	AGGAAGCCTG	2460
CCAAGGCGGC	TTGTTTCAGAG	AAATTGCTCC	TTCTGGTTTA	TGTCCAGCCT	TGATAACACA	2520
TATGGGAGCC	TACTATGCAG	TTTTAAAGCA	AGTATCCATG	CAGCCTGCAG	CCTGGTCATT	2580
TTTTCTGGGG	TGAGGATCTG	CCTAGGTAGA	AGTTTTCTCT	AATTTATTTT	GCTGTTACTT	2640
GTTATTGCAG	ATGGTTTCCT	GTCGGGGTGG	GGGTTTATT	TGCTTCCCAA	TGCTTTTGTT	2700
AATCCCGGTG	CTGTGTCTTA	TGTTGCAGTG	GTGGTGGTTC	TGGCATTAT	AATTTGCTGG	2760
TTGCCCTTCC	ACGTTGGCAG	AATCATTTAC	ATAAACACGG	AAGATTCGCG	GATGATGTAC	2820
TTCTCTCAGT	ACTTTAACAT	CGTCGCTAG	CAACTTTTCT	ATCTGAGCGC	ATCTATCAAC	2880
CCAATCCCT	ACAACCTCAT	TTCAAAGAAG	TACAGAGCGG	CGGCCTTTAA	ACTGCTGCTC	2940
GCAAGGAAGT	CCAGGCCGAG	AGGCTTCCAC	AGAAGCAGGG	ACACTGCGGG	GGAAGTTGCA	3000
GGGGACACTG	GAGGAGACAC	GGTGGGCTAC	ACCGAGACAA	GCGCTAACGT	GAAGACGATG	3060
GGATAA						3066

WO 99/64436

PCT/US99/12773

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1239 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

ATGGGCAGCC CCTGGAACGG CAGCGACGGC CCCGAGGGGG CGCGGGAGCC GCCGTGGCCC      60
GCGCTGCCCG CTTGCGACGA GCGCCGCTGC TCGCCCTTTC CCCTGGGGGG GCTGGTGCCC      120
GTGACCGCTG TGTGCCTGTG CCTGTTTCGTG GTCGGGGTGA GCGGCAACGT GGTGACCGTG      180
ATGCTGATCG GCGCTACCG GGACATGCGG ACCACCACCA ACTTGACCT GGGCAGCATG      240
GCCGTGTCCG ACCTACTCAT CCTGCTCGGG CTGCCGTTTC ACCTGTACCG CCTCTGGCGC      300
TCGCGGCCCT GGGTGTTCGG GCGCTGCTC TGCCGCCTGT CCCTCTACGT GGGCGAGGGC      360
TGCACCTACG CCACGCTGCT GCACATGACC GCGCTCAGCG TCGAGCGCTA CCTGGCCATC      420
TGCCGCCCCG TCCGCGCCCC CGTCTTGGTC ACCCGGCGCC GCGTCCGCGC GCTCATCGCT      480
GTGCTCTGGG CCGTGCGGCT GCTCTCTGCC GGTCCCTTCT TGTTCTTGGT GGGCGTCGAG      540
CAGGACCCCG GCATCTCCGT AGTCCCAGGC CTCAATGGCA CCGCGCGGAT CGCCTCCTCG      600
CCTCTCGCCT CGTCGCGGCC TCTCTGGCTC TCGCGGGCGC CACCGCCGTC CCCGCCGTCG      660
GGGCCCCGAGA CCGCGGAGGC CGCGGCGCTG TTCAGCCGCG AATGCCGGCC GAGCCCCGCG      720
CAGCTGGGCG CGCTGCGTGT CATGCTGTGG GTCACCACCG CCTACTTCTT CCTGCCCTTT      780
CTGTGCCTCA GCATCCTCTA CGGGCTCATC GGGCGGGAGC TGTGGAGCAG CCGGCGGCCG      840
CTGCGAGGCC CGGCCGCCTC GGGGCGGGAG AGAGGCCACC GGCAGACCGT CCGCGTCCTG      900
CTGGTGGTGG TTCTGGCATT TATAATTTGC TGGTTGCCCT TCCACGTTGG CAGAATCATT      960
TACATAAACA CGGAAGATTC GCGGATGATG TACTTCTCTC AGTACTTTAA CATCGTCGCT     1020
CTGCAACTTT TCTATCTGAG CGCATCTATC AACCCAATCC TCTACAACCT CATTTCAAAG     1080
AAGTACAGAG CGGCGGCCCT TAAACTGCTG CTCGCAAGGA AGTCCAGGCC GAGAGGCTTC     1140
CACAGAAGCA GGGACACTGC GGGGGAAGTT GCAGGGGACA CTGGAGGAGA CACGGTGGGC     1200
TACACCGAGA CAAGCGCTAA CGTGAAGACG ATGGGATAA                               1239

```

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 412 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Gly Ser Pro Trp Asn Gly Ser Asp Gly Pro Glu Gly Ala Arg Glu
 1           5           10           15
Pro Pro Trp Pro Ala Leu Pro Pro Cys Asp Glu Arg Arg Cys Ser Pro
 20           25           30
Phe Pro Leu Gly Ala Leu Val Pro Val Thr Ala Val Cys Leu Cys Leu
 35           40           45
Phe Val Val Gly Val Ser Gly Asn Val Val Thr Val Met Leu Ile Gly
 50           55           60
Arg Tyr Arg Asp Met Arg Thr Thr Thr Asn Leu Tyr Leu Gly Ser Met
 65           70           75           80

```

WO 99/64436

PCT/US99/12773

Ala	Val	Ser	Asp	Leu	Leu	Ile	Leu	Leu	Gly	Leu	Pro	Phe	Asp	Leu	Tyr
				85					90					95	
Arg	Leu	Trp	Arg	Ser	Arg	Pro	Trp	Val	Phe	Gly	Pro	Leu	Leu	Cys	Arg
			100					105					110		
Leu	Ser	Leu	Tyr	Val	Gly	Glu	Gly	Cys	Thr	Tyr	Ala	Thr	Leu	Leu	His
		115					120					125			
Met	Thr	Ala	Leu	Ser	Val	Glu	Arg	Tyr	Leu	Ala	Ile	Cys	Arg	Pro	Leu
	130					135					140				
Arg	Ala	Arg	Val	Leu	Val	Thr	Arg	Arg	Arg	Val	Arg	Ala	Leu	Ile	Ala
145					150					155				160	
Val	Leu	Trp	Ala	Val	Ala	Leu	Leu	Ser	Ala	Gly	Pro	Phe	Leu	Phe	Leu
			165						170					175	
Val	Gly	Val	Glu	Gln	Asp	Pro	Gly	Ile	Ser	Val	Val	Pro	Gly	Leu	Asn
			180					185					190		
Gly	Thr	Ala	Arg	Ile	Ala	Ser	Ser	Pro	Leu	Ala	Ser	Ser	Pro	Pro	Leu
		195					200					205			
Trp	Leu	Ser	Arg	Ala	Pro	Pro	Pro	Ser	Pro	Pro	Ser	Gly	Pro	Glu	Thr
	210					215					220				
Ala	Glu	Ala	Ala	Ala	Leu	Phe	Ser	Arg	Glu	Cys	Arg	Pro	Ser	Pro	Ala
225				230						235				240	
Gln	Leu	Gly	Ala	Leu	Arg	Val	Met	Leu	Trp	Val	Thr	Thr	Ala	Tyr	Phe
			245					250						255	
Phe	Leu	Pro	Phe	Leu	Cys	Leu	Ser	Ile	Leu	Tyr	Gly	Leu	Ile	Gly	Arg
			260					265					270		
Glu	Leu	Trp	Ser	Ser	Arg	Arg	Pro	Leu	Arg	Gly	Pro	Ala	Ala	Ser	Gly
		275					280					285			
Arg	Glu	Arg	Gly	His	Arg	Gln	Thr	Val	Arg	Val	Leu	Leu	Val	Val	Val
	290					295					300				
Leu	Ala	Phe	Ile	Ile	Cys	Trp	Leu	Pro	Phe	His	Val	Gly	Arg	Ile	Ile
305				310						315				320	
Tyr	Ile	Asn	Thr	Glu	Asp	Ser	Arg	Met	Met	Tyr	Phe	Ser	Gln	Tyr	Phe
			325					330						335	
Asn	Ile	Val	Ala	Leu	Gln	Leu	Phe	Tyr	Leu	Ser	Ala	Ser	Ile	Asn	Pro
		340					345					350			
Ile	Leu	Tyr	Asn	Leu	Ile	Ser	Lys	Lys	Tyr	Arg	Ala	Ala	Ala	Phe	Lys
		355					360					365			
Leu	Leu	Leu	Ala	Arg	Lys	Ser	Arg	Pro	Arg	Gly	Phe	His	Arg	Ser	Arg
	370					375					380				
Asp	Thr	Ala	Gly	Glu	Val	Ala	Gly	Asp	Thr	Gly	Gly	Asp	Thr	Val	Gly
385				390						395				400	
Tyr	Thr	Glu	Thr	Ser	Ala	Asn	Val	Lys	Thr	Met	Gly				
			405						410						

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1390 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGGGCAGCC CCTGGAACGG CAGCGACGGC CCCGAGGGGG CGCGGGAGCC GCCGTGGCCC

60

WO 99/64436

PCT/US99/12773

GCGCTGCCGC	CTTGCGACGA	GCGCCGCTGC	TCGCCCTTTC	CCCTGGGGGC	GCTGGTGCCG	120
GTGACCGCTG	TGTGCCTGTG	CCTGTTTCGT	GTCGGGGTGA	GCGGCAACGT	GGTGACCGTG	180
ATGCTGATCG	GGCGCTACCG	GGACATGCGG	ACCACCACCA	ACTTGTACCT	GGGCAGCATG	240
GCCGTGTCCG	ACCTACTCAT	CCTGCTCGGG	CTGCCGTTTC	ACCTGTACCG	CCTCTGGCGC	300
TCGCGGCCCT	GGGTGTTTCG	GCCGCTGCTC	TGCCCGCTGT	CCCTCTACGT	GGGCGAGGGC	360
TGCACCTACG	CCACGCTGCT	GCACATGACC	GCGCTCAGCG	TCGAGCGCTA	CCTGGCCATC	420
TGCCGCCCCG	TCCGCGCCCC	CGTCTTGGTC	ACCCGGCGCC	GCGTCCGCGC	GCTCATCGCT	480
GTGCTCTGGG	CCGTGGCGCT	GCTCTCTGCC	GGTCCCCTTC	TGTTCCCTGG	GGGCGTCGAG	540
CAGGACCCCG	GCATCTCCGT	AGTCCCGGGC	CTCAATGGCA	CCGCGCGGAT	CGCCTCCTCG	600
CCTCTCGCCT	CGTCGCCGCC	TCTCTGGCTC	TCGCGGGCGC	CACCGCCGTC	CCC GCCGTCG	660
GGGCCCCGAG	CCGCGGAGGC	CGCGGCGCTG	TTCAGCCGCG	AATGCCGGCC	GAGCCCCGCG	720
CAGCTGGGCG	CGCTGCGTGT	CATGCTGTGG	GTCACCACCG	CCTACTTCTT	CCTGCCCTTT	780
CTGTGCCTCA	GCATCCTCTA	CGGGCTCATC	GGGCGGGAGC	TGTGGAGCAG	CCGGCGGGCC	840
CTGCGAGGCC	CGGCCGCCCT	GGGCGGGAG	AGAGGCCACC	GGCAGACCGT	CCGCGTCCGT	900
CGTAAGTGA	GCCGCCGTGG	TTCCAAAGAG	GCTTGCCTGC	AGTCCGCCCC	GCCGGGGACC	960
GCGCAAACGC	TGGGTCCCCCT	TCCCCTGCTC	GCCCAGCTCT	GGGCGCCGCT	TCCAGCTCCC	1020
TTTCCTATTT	CGATTCCAGC	CTCCACCCGC	CGTGGTGGTG	GTTCTGGCAT	TTATAATTTG	1080
CTGGTTGCCC	TTCCACGTTG	GCAGAATCAT	TTACATAAAC	ACGGAAGATT	CGCGGATGAT	1140
GTACTTCTCT	CAGTACTTTA	ACATCGTCGC	TCTGCAACTT	TTCTATCTGA	GCGCATCTAT	1200
CAACCCAATC	CTCTACAACC	TCATTTCAA	GAAGTACAGA	GCGGCGGCCT	TTAAACTGCT	1260
GCTCGCAAGG	AAGTCCAGGC	CGAGAGGCTT	CCACAGAAGC	AGGGACACTG	CGGGGGAAGT	1320
TGCAGGGGAC	ACTGGAGGAG	ACACGGTGGG	CTACACCGAG	ACAAGCGCTA	ACGTGAAGAC	1380
GATGGGATAA						1390

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 386 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met	Gly	Ser	Pro	Trp	Asn	Gly	Ser	Asp	Gly	Pro	Glu	Gly	Ala	Arg	Glu
1				5					10					15	
Pro	Pro	Trp	Pro	Ala	Leu	Pro	Pro	Cys	Asp	Glu	Arg	Arg	Cys	Ser	Pro
			20					25					30		
Phe	Pro	Leu	Gly	Ala	Leu	Val	Pro	Val	Thr	Ala	Val	Cys	Leu	Cys	Leu
		35				40					45				
Phe	Val	Val	Gly	Val	Ser	Gly	Asn	Val	Val	Thr	Val	Met	Leu	Ile	Gly
	50					55					60				
Arg	Tyr	Arg	Asp	Met	Arg	Thr	Thr	Thr	Asn	Leu	Tyr	Leu	Gly	Ser	Met
65					70					75				80	
Ala	Val	Ser	Asp	Leu	Leu	Ile	Leu	Leu	Gly	Leu	Pro	Phe	Asp	Leu	Tyr
				85					90					95	
Arg	Leu	Trp	Arg	Ser	Arg	Pro	Trp	Val	Phe	Gly	Pro	Leu	Leu	Cys	Arg
			100					105					110		
Leu	Ser	Leu	Tyr	Val	Gly	Glu	Gly	Cys	Thr	Tyr	Ala	Thr	Leu	Leu	His
		115					120					125			
Met	Thr	Ala	Leu	Ser	Val	Glu	Arg	Tyr	Leu	Ala	Ile	Cys	Arg	Pro	Leu
	130					135					140				
Arg	Ala	Arg	Val	Leu	Val	Thr	Arg	Arg	Arg	Val	Arg	Ala	Leu	Ile	Ala
145					150					155					160

WO 99/64436

PCT/US99/12773

Val Leu Trp Ala Val Ala Leu Leu Ser Ala Gly Pro Phe Leu Phe Leu  
 165 170 175  
 Val Gly Val Glu Gln Asp Pro Gly Ile Ser Val Val Pro Gly Leu Asn  
 180 185 190  
 Gly Thr Ala Arg Ile Ala Ser Ser Pro Leu Ala Ser Ser Pro Pro Leu  
 195 200 205  
 Trp Leu Ser Arg Ala Pro Pro Ser Pro Pro Ser Gly Pro Glu Thr  
 210 215 220  
 Ala Glu Ala Ala Ala Leu Phe Ser Arg Glu Cys Arg Pro Ser Pro Ala  
 225 230 235 240  
 Gln Leu Gly Ala Leu Arg Val Met Leu Trp Val Thr Thr Ala Tyr Phe  
 245 250 255  
 Phe Leu Pro Phe Leu Cys Leu Ser Ile Leu Tyr Gly Leu Ile Gly Arg  
 260 265 270  
 Glu Leu Trp Ser Ser Arg Arg Pro Leu Arg Gly Pro Ala Ala Ser Gly  
 275 280 285  
 Arg Glu Arg Gly His Arg Gln Thr Val Arg Val Leu Arg Lys Trp Ser  
 290 295 300  
 Arg Arg Gly Ser Lys Asp Ala Cys Leu Gln Ser Ala Pro Pro Gly Thr  
 305 310 315 320  
 Ala Gln Thr Leu Gly Pro Leu Pro Leu Leu Ala Gln Leu Trp Ala Pro  
 325 330 335  
 Leu Pro Ala Pro Phe Pro Ile Ser Ile Pro Ala Ser Thr Arg Arg Gly  
 340 345 350  
 Gly Gly Ser Gly Ile Tyr Asn Leu Leu Val Ala Leu Pro Arg Trp Gln  
 355 360 365  
 Asn His Leu His Lys His Gly Arg Phe Ala Asp Asp Val Leu Leu Ser  
 370 375 380  
 Val Leu  
 385

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1092 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ATGCCCTGGA	CCAGACCCCA	GGTGACCTC	CATGCTGCTG	CAGCAGAGAC	CATGGACCAG	60
TACACCACGG	ACGACCACCA	CTACGAGGGC	TCCCTCTTCC	CCGCGTCCAC	CCTCATCCCC	120
GTCACGGTCA	TCTGCATCCT	CATCTTCGTG	GTCGGCGTGA	CCGGCAACAC	CATGACCATC	180
CTCATCATCC	AGTACTTCAA	GGACATGAAG	ACCACCACCA	ACCTGTACCT	GTCCAGCATG	240
GCCGTGTCCG	ACCTCGTCAT	CTTCCTCTGC	CTGCCCTTCG	ACCTGTACCG	CCTGTGGAAG	300
TACGTGCCGT	GGCTGTTCGG	CGAGGCCGTG	TGCCGCCTCT	ACCACTACAT	CTTCGAAGGC	360
TGCACGTCGG	CCACCATCCT	CCACATCAGC	GCCCTGAGCA	TCGAGCGCTA	CCTGGCCATC	420
AGCTTCCCCC	TCAGGAGCAA	GGTGATGGTG	ACCAGGAGAA	GGGTCCAGTA	CATCATCCTG	480
GCCCTGTGGT	GCTTCGCCCT	GGTGTGCGCC	GCTCCCACGC	TCTTCCTGGT	CGGGGTGGAG	540
TACGACAACG	AGACGCACCC	CGACTACAAC	ACGGGCCAGT	GCAAGCACAC	GGGCTACGCC	600
ATCAGCTCGG	GGCAGCTGCA	CATCATGATC	TGGGTGTCCA	CCACCTACTT	CTTCTGCCCG	660
ATGCTGTGTC	TCCTCTTCCT	CTACGGCTCC	ATCGGGTGCA	AGCTGTGGAA	GAGCAAGAAC	720
GACCTGCAGG	GCCCGTGCGC	CCTGGCCCCG	GAGAGGTCGC	ACAGGCAAAC	GGTGAAGATC	780

WO 99/64436

PCT/US99/12773

```

CTGGTGGTGG TGGTGCTGGC CTTCATCATC TGCTGGCTGC CCTACCACAT CGGCAGGAAC      840
CTGTTTCGCCC AGGTGGACGA CTACGACACG GCCATGCTCA GCCAGAATTT CAACATGGCC      900
TCCATGGTGC TCTGCTACCT CAGCGCCTCC ATCAACCCCG TCGTCTACAA CCTGATGTCG      960
AGGAAGTACC GGGCCGCCGC CAAGCGCCTC TTCCTGCTCC ACCAGAGACC CAAGCCGGCC     1020
CACCGGGGGC AGGGGCAGTT TTGCATGATC GGCCACAGCC CCACCCTGGA CGAGAGCCTG     1080
ACGGGGGTGT GA                                     1092

```

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 363 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```

Met Pro Trp Thr Arg Pro Gln Val Asp Leu His Ala Ala Ala Ala Glu
1      5      10      15
Thr Met Asp Gln Tyr Thr Thr Asp Asp His His Tyr Glu Gly Ser Leu
20      25      30
Phe Pro Ala Ser Thr Leu Ile Pro Val Thr Val Ile Cys Ile Leu Ile
35      40      45
Phe Val Val Gly Val Thr Gly Asn Thr Met Thr Ile Leu Ile Ile Gln
50      55      60
Tyr Phe Lys Asp Met Lys Thr Thr Thr Asn Leu Tyr Leu Ser Ser Met
65      70      75      80
Ala Val Ser Asp Leu Val Ile Phe Leu Cys Leu Pro Phe Asp Leu Tyr
85      90      95
Arg Leu Trp Lys Tyr Val Pro Trp Leu Phe Gly Glu Ala Val Cys Arg
100     105     110
Leu Tyr His Tyr Ile Phe Glu Gly Cys Thr Ser Ala Thr Ile Leu His
115     120     125
Ile Thr Ala Leu Ser Ile Glu Arg Tyr Leu Ala Ile Ser Phe Pro Leu
130     135     140
Arg Ser Lys Val Met Val Thr Arg Arg Arg Val Gln Tyr Ile Ile Leu
145     150     155     160
Ala Leu Trp Cys Phe Ala Leu Val Ser Ala Ala Pro Thr Leu Phe Leu
165     170     175
Val Gly Val Glu Tyr Asp Asn Glu Thr His Pro Asp Tyr Asn Thr Gly
180     185     190
Gln Cys Lys His Thr Gly Tyr Ala Ile Ser Ser Gly Gln Leu His Ile
195     200     205
Met Ile Trp Val Ser Thr Thr Tyr Phe Phe Cys Pro Met Leu Cys Leu
210     215     220
Leu Phe Leu Tyr Gly Ser Ile Gly Cys Lys Leu Trp Lys Ser Lys Asn
225     230     235     240
Asp Leu Gln Gly Pro Cys Ala Leu Ala Arg Glu Arg Ser His Arg Gln
245     250     255
Thr Val Lys Ile Leu Val Val Val Val Leu Ala Phe Ile Ile Cys Trp
260     265     270
Leu Pro Tyr His Ile Gly Arg Asn Leu Phe Ala Gln Val Asp Asp Tyr
275     280     285

```

WO 99/64436

PCT/US99/12773

Asp	Thr	Ala	Met	Leu	Ser	Gln	Asn	Phe	Asn	Met	Ala	Ser	Met	Val	Leu
290						295					300				
Cys	Tyr	Leu	Ser	Ala	Ser	Ile	Asn	Pro	Val	Val	Tyr	Asn	Leu	Met	Ser
305					310					315					320
Arg	Lys	Tyr	Arg	Ala	Ala	Lys	Arg	Leu	Phe	Leu	Leu	His	Gln	Arg	
			325					330					335		
Pro	Lys	Pro	Ala	His	Arg	Gly	Gln	Gly	Gln	Phe	Cys	Met	Ile	Gly	His
			340				345						350		
Ser	Pro	Thr	Leu	Asp	Glu	Ser	Leu	Thr	Gly	Val					
			355				360								

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCATCCTAAT ACGACTCACT ATAGGGC

27

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTATCCCATC GTCTTCACGT TAGCGCTTGT CTC

33

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTGCCCTTTC TGTGCCTCAG CATCCTCTAC

30

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:



WO 99/64436

PCT/US99/12773

- (A) LENGTH: 900 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

ATGGGCAGCC CCTGGAACGG CAGCGACGGC CCCGAGGGGG CGCGGGAGCC GCCGTGGCCC      60
GCGCTGCCGC CTTGCGACGA GCGCCGCTGC TCGCCCTTTC CCCTGGGGGC GCTGGTGCCG      120
GTGACCGCTG TGTGCCTGTG CCTGTTCGTC GTCGGGGTGA GCGGCAACGT GGTGACCGTG      180
ATGCTGATCG GCGGCTACCG GGACATGCGG ACCACCACCA ACTTGTACCT GGGCAGCATG      240
GCCGTGTCCG ACCTACTCAT CCTGCTCGGG CTGCCGTTTC ACCTGTACCG CCTCTGGCGC      300
TCGCGGCCCT GGGTGTTCGG GCCGCTGCTC TGCCGCTGT CCCTCTACGT GGGCGAGGGC      360
TGCACCTACG CCACGCTGCT GCACATGACC GCGCTCAGCG TCGAGCGCTA CCTGGCCATC      420
TGCCGCCCGC TCCGCGCCCG CGTCTTGGTC ACCCGGCGCC GCGTCCGCGC GCTCATCGCT      480
GTGCTCTGGG CCGTGGCGCT GCTCTCTGCC GGTCCCTTCT TGTTCCTGGT GGGCGTCGAG      540
CAGGACCCCG GCATCTCCGT AGTCCCGGGC CTCAATGGCA CCGCGCGGAT CGCCTCCTCG      600
CCTCTCGCCT CGTCGCGGCC TCTCTGGCTC TCGCGGGCGC CACCGCCGTC CCCGCCGTCG      660
GGGCCCCGAG CCGCGGAGGC CGCGGCGCTG TTCAGCCGCG AATGCCGGCC GAGCCCCGCG      720
CAGCTGGGCG CGCTGCGTGT CATGCTGTGG GTCACCACCG CCTACTTCTT CCTGCCCTTT      780
CTGTGCCTCA GCATCCTCTA CGGGCTCATC GGGCGGGAGC TGTGGAGCAG CCGGCGGCCG      840
CTGCGAGGCC CGGCCGCCTC GGGGCGGGAG AGAGGCCACC GGCAGACCGT CCGCGTCTGT      900
  
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 300 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```

Met Gly Ser Pro Trp Asn Gly Ser Asp Gly Pro Glu Gly Ala Arg Glu
 1          5          10          15
Pro Pro Trp Pro Ala Leu Pro Pro Cys Asp Glu Arg Arg Cys Ser Pro
 20          25          30
Phe Pro Leu Gly Ala Leu Val Pro Val Thr Ala Val Cys Leu Cys Leu
 35          40          45
Phe Val Val Gly Val Ser Gly Asn Val Val Thr Val Met Leu Ile Gly
 50          55          60
Arg Tyr Arg Asp Met Arg Thr Thr Thr Asn Leu Tyr Leu Gly Ser Met
 65          70          75          80
Ala Val Ser Asp Leu Leu Ile Leu Leu Gly Leu Pro Phe Asp Leu Tyr
 85          90          95
Arg Leu Trp Arg Ser Arg Pro Trp Val Phe Gly Pro Leu Leu Cys Arg
100          105          110
Leu Ser Leu Tyr Val Gly Glu Gly Cys Thr Tyr Ala Thr Leu Leu His
115          120          125
Met Thr Ala Leu Ser Val Glu Arg Tyr Leu Ala Ile Cys Arg Pro Leu
130          135          140
  
```

WO 99/64436

PCT/US99/12773

```

Arg Ala Arg Val Leu Val Thr Arg Arg Arg Val Arg Ala Leu Ile Ala
145          150          155          160
Val Leu Trp Ala Val Ala Leu Leu Ser Ala Gly Pro Phe Leu Phe Leu
          165          170          175
Val Gly Val Glu Gln Asp Pro Gly Ile Ser Val Val Pro Gly Leu Asn
          180          185          190
Gly Thr Ala Arg Ile Ala Ser Ser Pro Leu Ala Ser Ser Pro Pro Leu
          195          200          205
Trp Leu Ser Arg Ala Pro Pro Pro Ser Pro Pro Ser Gly Pro Glu Thr
          210          215          220
Ala Glu Ala Ala Ala Leu Phe Ser Arg Glu Cys Arg Pro Ser Pro Ala
225          230          235          240
Gln Leu Gly Ala Leu Arg Val Met Leu Trp Val Thr Thr Ala Tyr Phe
          245          250          255
Phe Leu Pro Phe Leu Cys Leu Ser Ile Leu Tyr Gly Leu Ile Gly Arg
          260          265          270
Glu Leu Trp Ser Ser Arg Arg Pro Leu Arg Gly Pro Ala Ala Ser Gly
          275          280          285
Arg Glu Arg Gly His Arg Gln Thr Val Arg Val Leu
          290          295          300

```

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 154 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```

CGTAAGTGGG  GCCGCCGTGG  TTCCAAAGAC  GCCTGCCTGC  AGTCCGCCCC  GCCGGGGACC  60
GCGCAAACGC  TGGGTCCCCT  TCCCCTGCTC  GCCCAGCTCT  GGGCGCCGCT  TCCAGCTCCC  120
TTTCCTATTT  CGATTCCAGC  CTCCACCCGC  CGGT                154

```

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 602 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: Genomic DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```

AGCTGGTGGT  GGTTCCTGGCA  TTTATAATTT  GCTGGTTGCC  CTTCCACGTT  GGCAGAATCA  60
TTTACATAAA  CACGGAAGAT  TCGCGGATGA  TGTACTTCTC  TCAGTACTTT  AACATCGTCG  120
CTCTGCAACT  TTTCTATCTG  AGCGCATCTA  TCAACCCAAT  CCTCTACAAC  CTCATTTCAA  180
AGAAGTACAG  AGCGGCGGCC  TTTAAACTGC  TGCTCGCAAG  GAAGTCCAGG  CCGAGAGGCT  240
TCCACAGAAG  CAGGGACACT  GCGGGGGAAG  TTGCAGGGGA  CACTGGAGGA  GACACGGTGG  300
GCTACACCGA  GACAAGCGCT  AACGTGAAGA  CGATGGGATA  ACGTAAGTGG  AGCCGCCGTG  360
GTTCCAAAGA  CGCCTGCCTG  CAGTCCGCCC  CGCCGGGGAC  CGCGCAAACG  CTGGGTCCCC  420

```

WO 99/64436

PCT/US99/12773

TTCCCCCTGCT	CGCCCAGCTC	TGGGCGCCGC	TTCCAGCTCC	CTTTCCTATT	TCGATTCCAG	480
CCTCCACCCG	CCGTGGTGGT	GGTTCTGGCA	TTTATAATTT	GCTGGTTGCC	CTTCCACGTT	540
GGCAGAATCA	TTTACATAAA	CACGGAAGAT	TCGCGGATGA	TGTACTTCTC	TCAGTACTTT	600
AA						602

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 198 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Leu	Val	Val	Val	Leu	Ala	Phe	Ile	Ile	Cys	Trp	Leu	Pro	Phe	His	Val
1				5					10					15	
Gly	Arg	Ile	Ile	Tyr	Ile	Asn	Thr	Glu	Asp	Ser	Arg	Met	Met	Tyr	Phe
			20					25					30		
Ser	Gln	Tyr	Phe	Asn	Ile	Val	Ala	Leu	Gln	Leu	Phe	Tyr	Leu	Ser	Ala
		35				40					45				
Ser	Ile	Asn	Pro	Ile	Leu	Tyr	Asn	Leu	Ile	Ser	Lys	Lys	Tyr	Arg	Ala
	50				55					60					
Ala	Ala	Phe	Lys	Leu	Leu	Leu	Ala	Arg	Lys	Ser	Arg	Pro	Arg	Gly	Phe
65					70				75					80	
His	Arg	Ser	Arg	Asp	Thr	Ala	Gly	Glu	Val	Ala	Gly	Asp	Thr	Gly	Gly
			85					90					95		
Asp	Thr	Val	Gly	Tyr	Thr	Glu	Thr	Ser	Ala	Asn	Val	Lys	Thr	Met	Gly
		100						105					110		
Arg	Lys	Trp	Ser	Arg	Arg	Gly	Ser	Lys	Asp	Ala	Cys	Leu	Gln	Ser	Ala
		115				120						125			
Pro	Pro	Gly	Thr	Ala	Gln	Thr	Leu	Gly	Pro	Leu	Pro	Leu	Leu	Ala	Gln
	130				135						140				
Leu	Trp	Ala	Pro	Leu	Pro	Ala	Pro	Phe	Pro	Ile	Ser	Ile	Pro	Ala	Ser
145					150				155					160	
Thr	Arg	Arg	Gly	Gly	Gly	Ser	Gly	Ile	Tyr	Asn	Leu	Leu	Val	Ala	Leu
			165					170					175		
Pro	Arg	Trp	Gln	Asn	His	Leu	His	Lys	His	Gly	Arg	Phe	Ala	Asp	Asp
			180					185					190		
Val	Leu	Leu	Ser	Val	Leu										
		195													